Smooth Solvation Method for d-Orbital Semiempirical Calculations of Biological Reactions. 1. Implementation

Jana Khandogin,[†] Brent A. Gregersen,[†] Walter Thiel,[‡] and Darrin M. York*[,†]

Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, and Max-Planck-Institut für Kohlenforschung, 45466 Mülheim an der Ruhr, Germany

Received: December 31, 2004; In Final Form: March 8, 2005

The present paper describes the extension of a recently developed smooth conductor-like screening model for solvation to a d-orbital semiempirical framework (MNDO/d-SCOSMO) with analytic gradients that can be used for geometry optimizations, transition state searches, and molecular dynamics simulations. The methodology is tested on the potential energy surfaces for separating ions and the dissociative phosphoryl transfer mechanism of methyl phosphate. The convergence behavior of the smooth COSMO method with respect to discretization level is examined and the numerical stability of the energy and gradient are compared to that from conventional COSMO calculations. The present method is further tested in applications to energy minimum and transition state geometry optimizations of neutral and charged metaphosphates, phosphates, and phosphoranes that are models for stationary points in transphosphorylation reaction pathways of enzymes and ribozymes. The results indicate that the smooth COSMO method greatly enhances the stability of quantum mechanical geometry optimization and transition state search calculations that would routinely fail with conventional solvation methods. The present MNDO/d-SCOSMO method has considerable computational advantages over hybrid quantum mechanical/molecular mechanical methods with explicit solvation, and represents a potentially useful tool in the arsenal of multi-scale quantum models used to study biochemical reactions.

1. Introduction

The importance of solvation effects on molecular dynamics and chemical reactions has motivated two decades of intensive effort in the development of solvent models for use in molecular dynamics simulations[1−4] and electronic structure calculations.[5,6] Implicit solvation models offer a great computational advantage over explicit models due to the tremendous reduction of the degrees of freedom that require force evaluation and configurational sampling. The most widely used implicit models include the finite-difference Poisson−Boltzmann (PB) method,[7−9] the generalized Born (GB) model,[1,3,10,11] and the boundary element method (BEM).[12−15]

One of the most powerful applications of implicit solvation methods involves their use with electronic structure calculations. In recent years, electronic structure methods with implicit solvation models have been applied to the prediction of approximate pK_a's, solution-phase electronic, NMR and EPR spectroscopic parameters, conformational equilibria, and chemical reaction kinetics.[5,6] Solvation models based on boundary element methods have been among the most widely applied in electronic structure calculations.[5] However, applications of these methods to potential energy surfaces (PES) and chemical reactions have traditionally been plagued by the lack of rigorously smooth analytic gradients that prevent reliable geometry optimization procedures. As a result, the studies of solvent effects on chemical reactions often default to solvation energy post correction based on gas-phase optimized geometries. The need for smooth analytic gradients is amplified for biological macromolecules where the degrees of freedom increase and conformational variations may involve large changes in the solvent exposed area. Recent advances in linear-scaling algorithms for electronic structure calculations[16−18] and boundary element solvation methods[19,21,22] have made possible application to very large systems such as solvated biological macromolecules at the semiempirical level.[23−25] These methods hold considerable promise combined with the linear-scaling geometry optimization and transition state search methods[26] in the study of reactions catalyzed by enzymes and ribozymes.

Boundary element methods have particular advantage for dielectric problems that involve a molecular cavity of unit dielectric surrounded by a solvent (such as water) modeled as a high dielectric constant. For such a model, the solvent reaction field potential can be represented as the potential arising from a surface charge distribution that lies at the dielectric boundary. This allows attention to be focused on a two-dimensional surface problem as opposed to the one involving a large three-dimensional volume. The advantage is that numerically it is possible to converge the discretization of the boundary element surface in a BEM solvation calculation much more readily than to increase the resolution of a three-dimensional grid in a finite-difference PB calculation, although the development of adaptive finite-element mesh methods offers improvement for the latter.[27,28] Numerical convergence is especially important when combined with quantum methods where the self-consistent reaction field has a significant polarization effect on the electronic structure.

BEM based methods such as the polarizable continuum model (PCM)[12,29] and conductor-like screening model (COSMO)[14,30] involve discretization of the cavity surface into tesserae used to expand the solvent polarization density from which the reaction field potential is derived. A difficulty that arises from the surface discretization in these methods is to ensure continuity

* Corresponding author: Phone: (612) 624-8042. Fax: (612) 626-7541. E-mail: york@chem.umn.edu. Web site: http://riesling.chem.umn.edu.
[†] University of Minnesota.
[‡] Max-Planck-Institut für Kohlenforschung.

10.1021/jp044062d CCC: $30.25
Published on Web 04/20/2005

in the solvation energy and derivatives with respect to the nuclear coordinates. The lack of rigorous analytic derivatives of the solvation energy and reaction field potential, as implemented in many electronic structure packages, often leads to failure in gradient-based computations, such as molecular geometry optimizations and transition state searches, and numerical instabilities in molecular dynamics simulations. One solution to improve stability in the BEM based solvation energy and gradient is to derive rigorous derivatives for the surface tesserae.[31,32] However, these methods may still be unreliable if the tessellation procedure itself is not a smooth function of the nuclear coordinates. A recent reformulation of a surface tessellation procedure and associated gradients leads to nearly continuous potential energy surfaces.[33] Alternately, the smooth COSMO method developed by York and Karplus has rigorously derived smooth analytic derivatives,[15] although until now, the method has not been integrated with an electronic structure method.

The present work reports the implementation of the smooth COSMO method[15] including analytic gradients within a d-orbital semiempirical framework.[34,35] The method allows the calculation of smooth potential energy surfaces as well as stable energy minimization and vibrational analysis (computation of second derivatives via finite difference of the analytic gradients). The paper is organized as follows. In the theory and methods section, the smooth COSMO method is briefly reviewed followed by the discussions of the implementation of the analytic gradients and the integration within the d-orbital semiempirical framework. A correction for multiple-atom switching is also introduced into the original smooth COSMO formulation. After a description of the computational details, the combined MNDO/d-smooth COSMO (MNDO/d-SCOSMO) method is tested on several systems. First, it is validated on the potential energy surfaces for separating ions in order to demonstrate smoothness of the energy and gradient, and coincidence of the stationary points with the zero-gradient norm. It is then applied to the dissociative mechanism of phosphoryl transfer in methyl phosphate and compared to the results obtained from the conventional COSMO method. The method is also tested in energy minimum and transition state optimizations for neutral and charged metaphosphates, phosphates and phosphoranes that are models for stationary points in transphosphorylation reaction pathways of enzymes and ribozymes. The final section draws conclusions and projects directions of future research.

2. Theory and Methods

2.1. Smooth COSMO Method. In the special case of a charge distribution ρ contained within a closed cavity of uniform dielectric $\epsilon_1 = 1$ surrounded by a medium of uniform dielectric ϵ_2, the polarization density σ_{pol}[15] resides only at the dielectric boundary (i.e., the cavity surface). In the limit that the external dielectric goes to infinity (i.e., a conductor), a simple variational principle emerges for the total polarization surface charge density

$$\delta E_{pol}/\delta\sigma_{pol} = 0 \tag{1}$$

where

$$E_{pol}[\sigma_{pol}] = \frac{1}{2}\int\int \sigma_{pol}(\mathbf{r})G_0(\mathbf{r},\mathbf{r}')\sigma_{pol}(\mathbf{r}')\,d^3r\,d^3r' + \int\int \sigma_{pol}(\mathbf{r})G_0(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')\,d^3r\,d^3r' \tag{2}$$

where $G_0(\mathbf{r},\mathbf{r}')$ is the Green's function in vacuo (e.g., $1/|\mathbf{r}-\mathbf{r},\mathbf{r}'|$ under real-space boundary conditions), and the integrals can usually be calculated analytically with the appropriate basis.

The conductor variational principle of eq 1 is the basis for the conductor-like screening model (COSMO) first proposed by Klamt and Schüürmann[15] and later extended by others.[15,19,30,36−39] In the York and Karplus variant of the COSMO model[15] that has rigorously smooth gradients (smooth COSMO), the solvent polarization energy is written in the algebraic form as

$$E_{pol}[\boldsymbol{\sigma}] = \frac{1}{2}\boldsymbol{\sigma}^T\cdot\mathbf{A}\cdot\boldsymbol{\sigma} + \boldsymbol{\sigma}^T\cdot\mathbf{B}\cdot\mathbf{q} \tag{3}$$

where $\boldsymbol{\sigma}$ is a $M \times 1$ vector representing the polarization surface charge density (the "pol" subscript will henceforth be dropped for $\boldsymbol{\sigma}$), $\mathbf{q}$ is a $N \times 1$ vector representing the solute charge distribution (the symbol $\mathbf{q}$ is used to distinguish from the electron density ρ that will be discussed in the next section), and $\mathbf{A}$ and $\mathbf{B}$ are $M \times M$ and $M \times N$ matrixes representing the self-interaction of the surface charge vectors and the interaction of the surface charge and solute charge density, respectively. In the general case, where there is an internal dielectric ϵ_1 and external dielectric ϵ_2, the $\mathbf{A}$ matrix is scaled by a factor $1/f(\epsilon_1,\epsilon_2)$ in accord with Gauss' law for the total reaction field surface charge density

$$\mathbf{A} = \frac{\mathbf{A}_0}{f(\epsilon_1,\epsilon_2)} \tag{4}$$

where $\mathbf{A}_0$ is the (unscaled) Coulomb interaction matrix and $f(\epsilon_1,\epsilon_2)$ is given by

$$f(\epsilon_1,\epsilon_2) = \frac{\epsilon_2 - \epsilon_1}{\epsilon_1\epsilon_2} \tag{5}$$

In the case of $\epsilon_1 \neq 1$, the polarization density includes a simple volume polarization term that takes the form of a scaled charge density, $\boldsymbol{\sigma}^V = \mathbf{q}(1-\epsilon_1)/\epsilon_1$.

In electronic structure calculations, the condition that the solute charge density is completely contained within the cavity is not strictly satisfied due to the exponential decay of the tail of the molecular electron density. Several remedies have been proposed to account for the effects due to outlying charges (also referred to as "charge penetration" effects).[15,40−42] One proposed solution involves the use of linear constraints on the total polarization surface charge in accord with Gauss' law for a conductor (including the possibility of higher order multipole moments) in the variational minimization of the solvent polarization energy. Accordingly, the constrained variational principle becomes

$$\delta\{E_{pol}[\boldsymbol{\sigma}] - \boldsymbol{\lambda}^T\cdot(\mathbf{D}^T\cdot\boldsymbol{\sigma} - \mathbf{Z}\cdot\mathbf{q})\} = 0 \tag{6}$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers on N_c constraint conditions

$$\mathbf{D}^T\cdot\boldsymbol{\sigma} = \mathbf{Z}\cdot\mathbf{q} \tag{7}$$

where $\mathbf{D}^T$ is a $N_c \times M$ matrix representing the N_c linear constraint equations and the vector $\mathbf{Z}\cdot\mathbf{q}$ contains the constraint values.

The surface charge vector obtained from the constrained variational minimization is

$$\boldsymbol{\sigma}^*(\boldsymbol{\lambda}) = -\mathbf{A}^{-1}(\mathbf{B}\cdot\mathbf{q} - \mathbf{D}\cdot\boldsymbol{\lambda}) = \boldsymbol{\sigma}^*(\mathbf{0}) + \delta\boldsymbol{\sigma}^*(\boldsymbol{\lambda}) \tag{8}$$

where $\boldsymbol{\sigma}^*(\mathbf{0}) = -\mathbf{A}^{-1}\cdot\mathbf{B}\cdot\mathbf{q}$ is the unconstrained surface charge vector and $\delta\boldsymbol{\sigma}^*(\boldsymbol{\lambda}) = \mathbf{A}^{-1}\cdot\mathbf{D}\cdot\boldsymbol{\lambda}$ is the constraint correction. The vector of Lagrange multipliers, $\boldsymbol{\lambda}$, is given by

$$\boldsymbol{\lambda} = \mathbf{Q}^{-1}\cdot\mathbf{R}\cdot\mathbf{q} \tag{9}$$

where $\mathbf{Q} = \mathbf{D}^{\mathrm{T}}\cdot\mathbf{A}^{-1}\cdot\mathbf{D}$ and $\mathbf{R} = (\mathbf{Z} + \mathbf{D}^{\mathrm{T}}\cdot\mathbf{A}^{-1}\cdot\mathbf{B})$. The resulting solvent polarization energy is then

$$\begin{aligned} E_{\mathrm{pol}}[\boldsymbol{\sigma}^*(\boldsymbol{\lambda})] &= E_{\mathrm{pol}}[\boldsymbol{\sigma}^*(\mathbf{0})] + \frac{1}{2}\,\delta\boldsymbol{\sigma}^*(\boldsymbol{\lambda})^{\mathrm{T}}\cdot\mathbf{A}\cdot\delta\boldsymbol{\sigma}^*(\boldsymbol{\lambda}) \\ &= \frac{1}{2}\,\mathbf{q}^{\mathrm{T}}\cdot[\mathbf{G}_{\mathrm{pol}}(\mathbf{0}) + \delta\mathbf{G}_{\mathrm{pol}}(\boldsymbol{\lambda})]\cdot\mathbf{q} \\ &= \frac{1}{2}\,\mathbf{q}^{\mathrm{T}}\cdot\mathbf{G}_{\mathrm{pol}}(\boldsymbol{\lambda})\cdot\mathbf{q} \end{aligned} \tag{10}$$

where $\mathbf{G}_{\mathrm{pol}}(\boldsymbol{\lambda}) = \mathbf{G}_{\mathrm{pol}}(\mathbf{0}) + \delta\mathbf{G}_{\mathrm{pol}}(\boldsymbol{\lambda})$ is the Green's function of the constrained variational procedure, $\mathbf{G}_{\mathrm{pol}}(\mathbf{0}) = -\mathbf{B}^{\mathrm{T}}\cdot\mathbf{A}^{-1}\cdot\mathbf{B}$ is the unconstrained Green's function, and $\delta\mathbf{G}_{\mathrm{pol}}(\boldsymbol{\lambda}) = \mathbf{R}^{\mathrm{T}}\cdot\mathbf{Q}^{-1}\cdot\mathbf{R}$ is the constraint correction. It is evident from the first line of eq 10 that the constrained variational energy is simply equal to the unconstrained variational energy, $E_{\mathrm{pol}}[\boldsymbol{\sigma}^*(\mathbf{0})]$, plus a positive semidefinite term that is the self-energy due to the constraint correction to the polarization density, $\delta\boldsymbol{\sigma}^*(\boldsymbol{\lambda})$. In the limit that the constraint correction vanishes, the unconstrained variational energy is recovered.

In the smooth COSMO method, the polarization surface charge density at the dielectric boundary used to model the solvent reaction field potential is described by a set of surface elements on a discretized solvent accessible surface. The discretized surface is constructed as a superposition of individually discretized atom-centered spheres (adjusted by a solvent probe radius). The discretization of each atomic sphere, described in detail elsewhere,[15] was derived from sets of points $\{\hat{\mathbf{r}}_k\}$ and weights $\{w_k\}$ used in high-order numerical angular quadrature schemes with octahedral symmetry adapted for integration of spherical harmonic functions,[43] first pioneered by Lebedev[44] and extended to high order by Delley.[45]

The smooth COSMO method uses spherical Gaussian functions of the form

$$g_k(|\mathbf{r} - \mathbf{r}_k|) = (\zeta_k{}^2/\pi)^{3/2}\mathrm{e}^{-\zeta_k{}^2|\mathbf{r}-\mathbf{r}_k|^2} \tag{11}$$

to represent the electrostatic potential and surface element interaction matrixes and circumvents the Coulomb singularity problem due to overlapping surface elements represented by point charges. Here, $\mathbf{r}_k$ is the coordinate of the surface element k and ζ_k is the Gaussian exponent that is adjusted to obtain the exact Born ion solvation energy.[15]

In order for the appearance and disappearance of surface elements to occur smoothly as a function of geometrical changes, a switching layer around each atom is introduced (Figure 1). The switching layer for jth atom has an inner radius (R_{in_j}) and outer radius (R_{out_j}) defined by

$$R_{\mathrm{in}_j} = R_j - \alpha_j R_{\mathrm{sw}_j} \tag{12}$$

$$R_{\mathrm{out}_j} = R_j + (1-\alpha_j)R_{\mathrm{sw}_j} \tag{13}$$

where R_j is the atomic radius (plus solvent probe radius) of atom j, R_{sw_j} is the switching layer thickness, and α_j is the switching layer shift parameter (see below). The switching layer serves to "turn off" or "turn on" the surface elements associated with other atoms as they pass into or out of the layer. This effect is

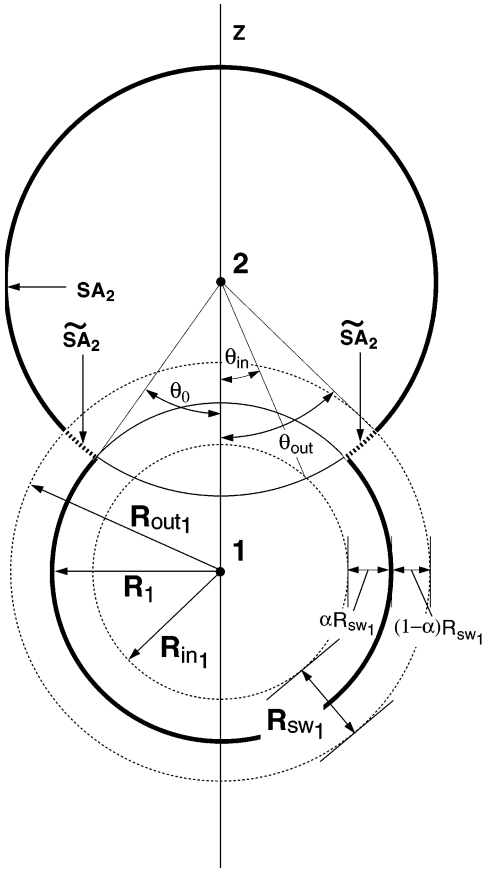

Figure 1. Switching layer in the smooth COSMO method that allows the appearance and disappearance of surface elements to occur smoothly as a function of geometrical changes.

brought about by scaling the surface element self-energy values from their calibrated values at R_{out_j} rapidly but smoothly to infinity as they reach R_{in_j}. In the present work, the switching function takes the form

$$S_k = [\prod_j^{\mathrm{atoms}} S_{\mathrm{wf}}(\hat{r}_{k,j})]^p \quad \text{for } k \notin j \tag{14}$$

where p is a parameter, not present in the original smooth COSMO formulation,[15] which accounts for surface elements that are, on average, being switched simultaneously by more than one other atom, and $S_{\mathrm{wf}}(\hat{r}_{k,j})$ is the switching function given by

$$S_{\mathrm{wf}}(r) = \begin{cases} 0 & r < 0 \\ r^3(10 - 15r + 6r^2) & 0 \le r \le 1 \\ 1 & r < 1 \end{cases} \tag{15}$$

and $\hat{r}_{k,j}$ is defined as

$$\hat{r}_{k,j} = [|\mathbf{r}_k - \mathbf{R}_j| - R_{\mathrm{in}_j}]/R_{\mathrm{sw}_j} \text{ for } k \notin j \tag{16}$$

Tests of different values of the p parameter for large molecules suggest that a value of around 0.25 may give more uniform convergence with respect to discretization level than the original formulation with p set to unity, although further testing is required.

In the original smooth COSMO formulation, the switching layer shift parameter in eq 12 was taken in accord with a simple analytic surface area formula for two overlapping atomic spheres

$$\alpha_j = \frac{1}{2} + \frac{R_j}{R_{sw_j}} - \sqrt{\left(\frac{R_j}{R_{sw_j}}\right)^2 - \beta^2} \tag{17}$$

where β is a parameter that depends on the form of the switching function and in the present case (eq 15) is equal to $(2\sqrt{7})^{-1}$. This choice of the α_j parameter is independent of the relative positions of the radii and, hence, does not require derivative terms in the formulas for the gradient.

The surface element interaction matrix $\mathbf{A}_0$ of eq 4 is given by

$$(\mathbf{A}_0)_{k,k} = \frac{2\zeta_k}{\sqrt{2\pi}} S_k^{-1} \tag{18}$$

$$(\mathbf{A}_0)_{k,l} = \frac{\mathrm{erf}(\zeta'_{k,l}|\mathbf{r}_k - \mathbf{r}_l|)}{|\mathbf{r}_k - \mathbf{r}_l|} \tag{19}$$

where k, l are indexes for the surface elements, S_k is the smooth switching parameter of eq 14, and $\zeta'_{k,l}$ is given by

$$\zeta'_{k,l} = \zeta_k\zeta_l/\sqrt{\zeta_k^2 + \zeta_l^2} \tag{20}$$

where ζ_k and ζ_l are the Gaussian exponents of the k and l surface elements, respectively.

The **B** matrix of eq 3 that describes the interaction between the reaction field surface Gaussian charge (σ) and solute charge distribution (**q**), in the case that the latter is a point charge distribution, takes the form

$$(\mathbf{B})_{k,j} = \frac{\mathrm{erf}(\zeta_k|\mathbf{r}_k - \mathbf{R}_j|)}{|\mathbf{r}_k - \mathbf{R}_j|} \tag{21}$$

where k is the surface element index and j is the index of the solute atomic point charge. In the next section, the formulas for the **B** matrix elements will be extended to the multipole interactions used in the d-orbital semiempirical formalism.

In addition to the electrostatic component, there are other terms that contribute to the solvation energy, including the energy required to form a cavity when placing a solute molecule in the solvent and the repulsion-dispersion interactions between the solute and solvent. These terms are frequently collected into a single term, referred to here as the nonpolar component of the solvation energy, E_{npol}. In this work, E_{npol} will take the general form[19,46,47]

$$E_{npol} = \sum_j^{atoms} \gamma_j \cdot SA_j \tag{22}$$

where γ_j is an effective microscopic surface tension parameter that takes into account cavity formation and dispersion−repulsion interactions[47,47] for atom j, and SA_j is the surface area on the solvent accessible surface associated with atom j given by

$$SA_j = \sum_{k\in j} w_k S_k \tag{23}$$

where the summation over k is over all surface elements associated with atom j, w_k is the angular quadrature weight,[19] and S_k is the scale factor as given in eq 14. Modifications of the switching layer shift parameter in accord with this formula may lead to further improvements of the model but have not been explored in the present study. The use of modified forms of the switching function and switching layer shift parameter and their relation to the solvent accessible and solvent excluding surfaces and impact on the solvation energy will be investigated in future work.

The total smooth COSMO solvation energy, E_{sol}, is given by the sum of the solvent polarization and nonpolar contributions

$$E_{sol} = E_{pol} + E_{npol} \tag{24}$$

2.2. Implementation of Analytic Energy Gradients. This section discusses the implementation of the analytic gradient for the smooth COSMO method, with an emphasis on modifications of the original derivation.[15] Formulas will be restricted to the derivatives of the unconstrained variational form of the COSMO energy, i.e., $E_{pol}[\sigma^*(0)]$ in eq 10.

Consider a set of solute atoms with positions $\mathbf{R}_m$, where m is an atom index, described by a charge distribution vector **q**. The gradient of the unconstrained solvation energy, $E_{pol}[\sigma^*(0)]$ in eq 10, with respect to $\mathbf{R}_m$ is given as

$$\begin{aligned}\nabla_m E_{pol} &= \left[\frac{\delta E_{pol}}{\delta \sigma_{pol}}\right]_{\sigma^*_{pol}} \cdot \frac{\partial \sigma^*_{pol}}{\partial \mathbf{R}_m} + \left[\frac{\delta E_{pol}}{\delta \mathbf{q}}\right]\mathbf{q}_0 \cdot \frac{\partial \mathbf{q}}{\partial \mathbf{R}_m} \\ &+ \frac{1}{2}\sigma^{*\,T}_{pol} \cdot \frac{\partial \mathbf{A}}{\partial \mathbf{R}_m} \cdot \sigma^*_{pol} + \sigma^{*\,T}_{pol} \cdot \frac{\partial \mathbf{B}}{\partial \mathbf{R}_m} \cdot \mathbf{q} \\ &= \frac{1}{2}\sigma^{*\,T}_{pol} \cdot \nabla_m \mathbf{A} \cdot \sigma_{pol} + \sigma^{*\,T}_{pol} \cdot \nabla_m \mathbf{B} \cdot \mathbf{q} \\ &= \frac{1}{2}\mathbf{q}^T \cdot \nabla_m \mathbf{G}_{pol}(0) \cdot \mathbf{q}\end{aligned} \tag{25}$$

The first term that involves $\delta E_{pol}/\delta\sigma_{pol}$ vanishes in the case of the unconstrained variational solvation energy, and the second term that involves $\partial\mathbf{q}/\partial\mathbf{R}_m$ vanishes in the case the charge distribution vector **q** does not depend on the atomic positions (e.g., if **q** is a static charge distribution; for generalization to quantum charge distributions, see below). The last line of the above equation provides a compact Green's function solution that can be generalized to incorporate constraints by replacement of the $\mathbf{G}_{pol}(0)$ term by $\mathbf{G}_{pol}(\lambda)$ of eq 10.

The gradient for the diagonal elements of the $\mathbf{A}_0$ matrix (which would be zero in the absence of switching) is given by

$$\nabla_m (\mathbf{A}_0)_{k,k} = -(\mathbf{A}_0)_{k,k} S_k^{-2} \nabla_m S_k \tag{26}$$

where

$$\nabla_m S_k = p \cdot S_k \sum_j \frac{\partial S_{wf}(\hat{r}_{k,j})}{\partial \hat{r}_{k,j}} \nabla_m \hat{r}_{k,j} \tag{27}$$

$$\frac{\partial S_{wf}(r)}{\partial r} = \begin{cases} 0 & r < 0 \\ 30r^2(r-1)^2 & 0 \le r \le 1 \\ 0 & r < 1 \end{cases} \tag{28}$$

$$\nabla_m \hat{r}_{k,j} = \frac{\mathbf{r}_k - \mathbf{R}_j}{|\mathbf{r}_k - \mathbf{R}_j|} \frac{1}{R_{sw_j}} \quad \text{for } k \in m \tag{29}$$

$$= \frac{\mathbf{r}_k - \mathbf{R}_j}{|\mathbf{r}_k - \mathbf{R}_j|} \frac{1}{R_{sw_j}} \delta_{jm} \quad \text{for } k \notin m \tag{30}$$

The gradient for the off-diagonal elements of the **A** matrix is

$$\begin{aligned}&\nabla_m (\mathbf{A}_0)_{k,l} = \\ &\quad -\frac{\mathbf{r}_k - \mathbf{r}_l}{|\mathbf{r}_k - \mathbf{r}_l|^2}\left((\mathbf{A}_0)_{k,l} - \frac{2}{\sqrt{\pi}}\zeta'_{k,l} e^{-\zeta'^2_{k,l}|\mathbf{r}_k - \mathbf{r}_l|^2}\right)(\delta_{im} - \delta_{jm})\end{aligned} \tag{31}$$

where $k \in i$, and $l \in j$. The gradient for the **B** matrix is

$$\nabla_m(\mathbf{B})_{k,j} = -\frac{\mathbf{r}_k - \mathbf{R}_j}{|\mathbf{r}_k - \mathbf{R}_j|^2}\left((\mathbf{B})_{k,j} - \frac{2}{\sqrt{\pi}}\zeta_k e^{-\zeta_k^2|\mathbf{r}_k-\mathbf{R}_j|^2}\right)(\delta_{im} - \delta_{jm}) \quad (32)$$

The gradient of the nonpolar solvation energy can be written as

$$\nabla_m E_{\mathrm{npol}} = \sum_j^{\mathrm{atoms}} \gamma_j \sum_{k\in j} w_k \nabla_m S_k \quad (33)$$

where $\nabla_m S_k$ is given in eq 27.

2.3. Integration with a d-Orbital Semiempirical Quantum Method. For a fixed geometry, the total solution-phase energy of a molecule can be written as

$$E[\rho] = E_0[\rho] + E_{\mathrm{sol}}[\rho] + E_{\mathrm{nuc}} \quad (34)$$

where $E_0[\rho]$ is the functional that returns the total gas-phase electronic energy for the solvent-polarized electron density ρ, E_{nuc} is the nuclear−nuclear repulsion energy, and $E_{\mathrm{sol}}[\rho]$ is the solvation energy of eq 24. Note that E_{pol} in eq 24 is the only term in the solvation energy that depends explicitly on the electron density through the total charge distribution vector **q**.

Since the solvation energy can be written as a functional of ρ through the dependence on the charge distribution vector **q**, it must be included in the variational procedure of the molecular electronic energy, leading to the solution-phase Fock matrix

$$F_{\mu\nu} = \frac{\delta E[\rho]}{\delta \rho_{\mu\nu}} = F^0_{\mu\nu} + \delta F^{\mathrm{RF}}_{\mu\nu} \quad (35)$$

where $\rho_{\mu\nu}$ is the single-particle density matrix, $F^0_{\mu\nu} = \delta E_0[\rho]/\delta\rho_{\mu\nu}$ is the Fock matrix in the absence of solvation, and $\delta F^{\mathrm{RF}}_{\mu\nu} = \delta E_{\mathrm{sol}}[\rho]/\delta\rho_{\mu\nu}$ is the correction to the Fock matrix for the solvent reaction field potential (see below).

From eq 34, the gradient of the solution-phase molecular energy with respect to the nuclear coordinate is

$$\nabla_m E[\rho] = \nabla_m E_0[\rho] + \nabla_m E_{\mathrm{pol}}[\rho] + \nabla_m E_{\mathrm{npol}} + \nabla_m E_{\mathrm{nuc}} + \int\left[\frac{\delta(E_0[\rho]+E_{\mathrm{pol}}[\rho])}{\delta\rho(\mathbf{r})}\right]_{\rho 0} \nabla_m \rho_0(\mathbf{r})\, \mathrm{d}^3 r \quad (36)$$

where ρ_0 is the ground-state electron density in solution. The last term vanishes for molecules that are constrained to a fixed number of electrons, since

$$\int\left[\frac{\delta(E_0+E_{\mathrm{pol}})}{\delta\rho(\mathbf{r})}\right]_{\rho 0} \nabla_m \rho_0(\mathbf{r})\, \mathrm{d}^3 r = \mu \nabla_m \int \rho_0(\mathbf{r})\, \mathrm{d}^3 r = 0 \quad (37)$$

where μ is the chemical potential of the system in solution.

For semiempirical NDDO-type methods, the solute charge can be written in a vector form obtainable from the atom-block diagonal elements (one-center basis pairs) of the density matrix

$$\mathbf{q}_{i,\mu\nu} = \begin{cases} \mathbf{q}_i^{\mathrm{core}} - \rho_{i,\mu\nu} & \mu\nu = ss \\ -\rho_{i,\mu\nu} & \mu\nu \notin sp, sd, pd, pp, dd \end{cases} \quad (38)$$

where the double index $i,\mu\nu$ refers to the basis pair $\mu\nu$ centered on the ith atom, $\mathbf{q}_i^{\mathrm{core}}$ is the effective core charge (nuclear charge and core electrons) of that atom, and $\rho_{i,\mu\nu}$ is the density matrix element $\mu\nu$ of the ith atom block.

The **B** matrix that describes the Coulomb interaction between the solute and surface charges can be written as

$$\mathbf{B}_{k,(i,\mu\nu)} = \int \tilde{\mathbf{B}}_{(i,\mu\nu)}(\mathbf{r}) g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r \quad (39)$$

where the index $k,(i,\mu\nu)$ refers to the interaction of the kth surface element with the density matrix element $(i,\mu\nu)$ and $g_k(|\mathbf{r} - \mathbf{r}_k|)$ is the spherical Gaussian function. $\tilde{\mathbf{B}}_{(i,\mu\nu)}(\mathbf{r})$ represents the electrostatic potential at **r** arising from the density matrix element $(i,\mu\nu)$. In accord with the semiempirical NDDO formalism,[48] $\tilde{\mathbf{B}}_{(i,\mu\nu)}(\mathbf{r})$ can be approximated as the potential due to the atomic monopole, dipole, and quadrupole moments. With inclusion of d orbitals,[34] the $\tilde{\mathbf{B}}$ matrix contains the following nonzero elements:

basis	matrix element	$\tilde{\mathbf{B}}_{(i,\mu\nu)}(\mathbf{r}_k)$	
μ	ν		
s	s	$\frac{1}{\lvert r_{k,i}\rvert}$	(40)
s	p_l	$\frac{d_{2,i}(r_{k,i})_l}{\lvert r_{k,i}\rvert^3}$	(41)
p_l	$p_{l'}$	$\frac{1}{\lvert r_{k,i}\rvert}\delta_{ll'} - \frac{d_{3,i}}{\lvert r_{k,i}\rvert^3}\delta_{ll'} + \frac{3d_{3,i}(r_{k,i})_l^2}{\lvert r_{k,i}\rvert^5}$	(42)
s	$d_{l'}$	$\frac{3d_{4,i}(r_{k,i})_l(r_{k,i})_{l'}}{\lvert r_{k,i}\rvert^5}$	(43)
s	$d_{x^2-y^2}$	$\frac{3d_{4,i}(r_{k,i})_x^2 - (r_{k,i})_y^2}{2\lvert r_{k,i}\rvert^5}$	(44)
s	d_{z^2}	$\frac{\sqrt{3}d_{4,i}[2(r_{k,i})_z^2 - (r_{k,i})_x^2 - (r_{k,i})_y^2]}{2\lvert r_{k,i}\rvert^5}$	(45)
p_l	$d_{l'}$	$\frac{d_{5,i}(r_{k,i})_{l'}}{\lvert r_{k,i}\rvert^3}$, $l \neq l'$	(46)
p_l	$d_{x^2-y^2}$	$\frac{d_{5,i}}{\lvert r_{k,i}\rvert^3}[(r_{k,i})_x\delta_{lx} + (r_{k,i})_y\delta_{ly}]$	(47)
p_l	d_{z^2}	$\frac{d_{5,i}}{\sqrt{3}\lvert r_{k,i}\rvert^3}[-(r_{k,i})_x\delta_{lx} - (r_{k,i})_y\delta_{ly} + 2(r_{k,i})_z\delta_{lz}]$	(48)
d_{z^2}	d_{z^2}	$\frac{1}{\lvert r_{k,i}\rvert} - \frac{d_{6,i}}{\lvert r_{k,i}\rvert^3} + \frac{3d_{6,i}(r_{k,i})_z^2}{\lvert r_{k,i}\rvert^5}$	(49)
$d_{x^2-y^2}$	$d_{x^2-y^2}$	$\frac{1}{\lvert r_{k,i}\rvert} + \frac{d_{6,i}}{\lvert r_{k,i}\rvert^3} - \frac{3d_{6,i}(r_{k,i})_z^2}{\lvert r_{k,i}\rvert^5}$	(50)
$d_{l'}$	$d_{l'}$	$\frac{1}{\lvert r_{k,i}\rvert} + \frac{d_{6,i}}{\lvert r_{k,i}\rvert^3} - \frac{3d_{6,i}(r_{k,i})_{l''}^2}{\lvert r_{k,i}\rvert^5}$, $l \neq l' \neq l''$	(51)
$d_{l'}$	$d_{l''}$	$\frac{3d_{6,i}(r_{k,i})_l(r_{k,i})_{l''}}{\lvert r_{k,i}\rvert^5}$, $l \neq l' \neq l''$	(52)
$d_{l'}$	d_{z^2}	$\frac{\sqrt{3}d_{6,i}}{\lvert r_{k,i}\rvert^5}[(r_{k,i})_l(r_{k,i})_{l'}\delta_{lz} - 2(r_{k,i})_x(r_{k,i})_y\delta_{lx}\delta_{l'y}]$, $l \neq l'$	(53)
d_{lz}	$d_{x^2-y^2}$	$\frac{3d_{6,i}}{\lvert r_{k,i}\rvert^5}[(r_{k,i})_x(r_{k,i})_z\delta_{lx} - (r_{k,i})_y(r_{k,i})_z\delta_{ly} - ((r_{k,i})_x^2 - (r_{k,i})_y^2)\delta_{lz}]$	(54)

where l and l' represent the Cartesian components x, y, or z, $\mathbf{r}_{k,i} = \mathbf{r}_k - \mathbf{R}_i$, and $d_{2,i}$, $d_{3,i}$, $d_{4,i}$, $d_{5,i}$, and $d_{6,i}$ are the multipole charge separations corresponding to *sp*, *pp*, *sd*, *pd*, and *dd* basis pairs centered on the ith atom, respectively.[34]

In light of eqs 39 and 40, the **B** matrix involves the Gaussian-monopole integral

$$I_0 = \int \frac{1}{|\mathbf{R}_i - \mathbf{r}|} g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r = \frac{\mathrm{erf}(\zeta_k |\mathbf{R}_i - \mathbf{r}_k|)}{|\mathbf{R}_i - \mathbf{r}_k|} \quad (55)$$

and the following Gaussian-multipole integrals

$$I_1 = \int \frac{1}{|\mathbf{R}_i - \mathbf{r}|^3} g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r$$

$$I_2 = \int \frac{1}{|\mathbf{R}_i - \mathbf{r}|^5} g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r$$

$$I_{1,l} = \int \frac{\tilde{r}_l}{|\mathbf{R}_i - \mathbf{r}|^3} g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r, \quad l = x, y, \text{ or } z$$

$$I_{2,ll'} = \int \frac{\tilde{r}_l \tilde{r}_{l'}}{|\mathbf{R}_i - \mathbf{r}|^5} g_k(|\mathbf{r} - \mathbf{r}_k|)\, \mathrm{d}^3 r, \quad l \neq l' \quad (56)$$

where $\tilde{\mathbf{r}} = \mathbf{R}_i - \mathbf{r}$. Notice that the Gaussian-monopole integral is consistent with eq 21. Since I_0 has an analytic form, the Gaussian-multipole integrals can be derived from the following differential relations:

$$\frac{d^2}{dr^2} I_0 = 2I_1$$

$$\frac{\partial}{\partial r_l} I_0 = -I_{1,l}$$

$$\frac{\partial}{\partial r_l} I_{1,l'} = -3I_{2,ll} + I_1 \delta_{l,l'} \quad (57)$$

Once $\mathbf{B}_{k,(i,\mu\nu)}$ is obtained, the gradient easily follows from eq 32. The value of $\mathbf{B}_{k,(i,\mu\nu)}$ is, in general, very close to that of $\hat{\mathbf{B}}_{(i,\mu\nu)}(\mathbf{r}_k)$, because the distance between the Gaussian function and atomic multipole is much larger than the width of the Gaussian function. The total reaction field potential that is added to the Fock matrix elements corresponding to the one-center basis pairs (see eq 35) can be calculated as

$$\delta F_{\mu\nu}^{\mathrm{RF}} = \sum_k \mathbf{B}_{k,(i,\mu\nu)} \sigma_k^* \quad (58)$$

where σ_k^* is the surface charge at $\mathbf{r}_k$ due to both effective core charge and valence electrons.

2.4. Computational Details. All test calculations were performed using the combined MNDO/d-smooth COSMO method implemented in MNDO97.[49] In the electronic structure calculations, the (default) tolerance to SCF convergence was set to 10^{-6} eV on the energy and 10^{-10} au on the norm of the density matrix. For the geometry optimizations of biological phosphorus compounds, the (default) tolerance for geometry convergence was 1 kcal/mol/Å on the gradient norm, whereas a more stringent tolerance of 0.01 kcal/mol/Å was used in the adiabatic reaction paths for dissociation of methyl phosphate to obtain smooth PESs.

Unless otherwise noted, the solvent accessible surface was constructed with the atomic radii due to Bondi,[50] with a surface discretization level of 110 points per sphere, and the parameters used in the smooth COSMO switching were $\gamma_s = 1.0$ and $\alpha \approx 0.5$ in accord with eq 17. In all of the calculations that follow, the dielectric model involves a fixed internal dielectric constant of $\epsilon_1 = 1$ (vacuo) in the region of the solute as defined by the solvent accessible surface, and a fixed external dielectric constant of $\epsilon_2 = 80$ (water) outside the solvent accessible surface (with the exception of NaCl where a constant external dielectric of $\epsilon_2 = 2$ was used, see below). The nonpolar component of the solvation energy (E_{npol} in eq 24) was neglected for two reasons. First, this term is small relative to the electrostatic counterpart for the ionic systems under study (typically less than 2%). For example, the electrostatic and nonelectrostatic solvation energy component predicted by density-functional calculations[51] are −66.4 and +0.3 kcal/mol, respectively, for a metaphosphate anion (PO_3^-). Second, the parameter does not vary excessively during geometry optimizations and/or along reaction pathways.

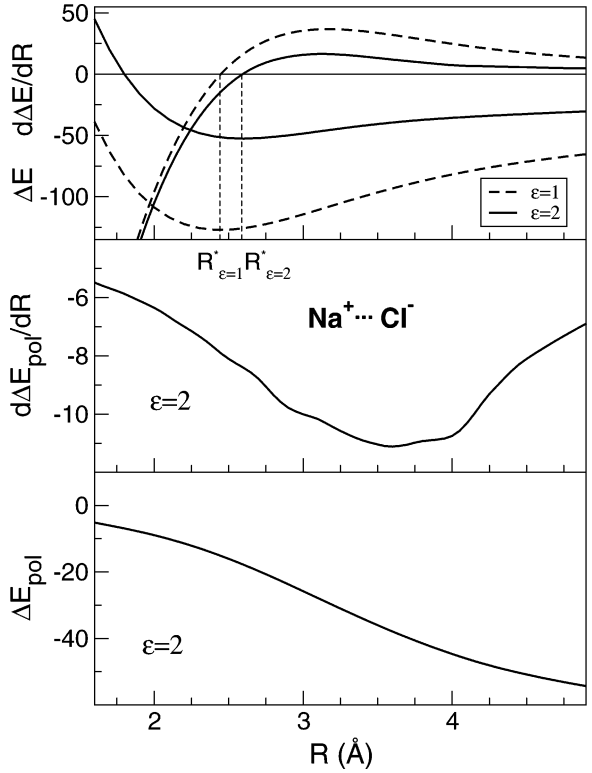

Figure 2. Binding energies (kcal/mol) and energy gradients (kcal/mol/Å) of NaCl as a function of internuclear distance R (Å). Top: binding energy (ΔE) and binding energy gradient ($\mathrm{d}\Delta E/\mathrm{d}R$). Dotted and solid black curves are from the gas ($\epsilon = 1$) and low-dielectric ($\epsilon = 2$) calculations, respectively. The stationary points (zero gradient) are indicated by R^* for each curve. The use of low dielectric of $\epsilon = 2$ was necessary for creating a minimum in the total binding energy along the ion separation coordinate (see text). Middle: electrostatic solvation component of the binding energy gradient ($\mathrm{d}\Delta E_{\mathrm{pol}}/\mathrm{d}R$). Bottom: electrostatic solvation component of the binding energy (ΔE_{pol}).

In the case of the separating $Na^+\cdots Cl^-$ ions, to create a minimum in the total energy curve, a low dielectric of $\epsilon = 2.0$ was used. The solvation radii for Na^+ and Cl^- were taken as the Born ion radii of 1.66 and 2.15 Å,[52] respectively. The smooth COSMO model for d-orbital semiempirical methods has been implemented into the MNDO code[49] and interfaced to the CHARMM molecular modeling package[53] and will be available in future MNDO and CHARMM releases.

3. Results and Discussion

3.1. Potential Energy Curve for NaCl. The purpose of this subsection is to examine a benchmark potential energy surface (PES) in order to test the smoothness of solvation energy and gradient and demonstrate the coincidence of the energy minimum and zero-gradient points.

Figure 2 (top) shows the gas-phase (dotted, $\epsilon = 1$) and low-dielectric (solid, $\epsilon = 2$) binding energies and total energy

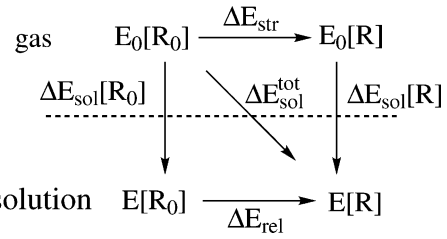

Figure 3. Thermodynamic cycle of the solvation process. E_0 and E refer to gas-phase and solution-phase energy calculations, respectively, and indicated in brackets are the geometries optimized in the gas phase (R_0) or in solution (R).

gradients for $Na^+\cdots Cl^-$ as a function of the internuclear distance. If the dielectric is set to that of water (roughly 80), the solvent stabilization produces an energy curve that decreases monotonically with internuclear separation (data not shown). Consequently, for demonstration purposes, a low dielectric ($\epsilon = 2$) was used to create a minimum for illustrating the coincidence of the minimum energy and zero-gradient distance values. As the total energy gradient approaches zero, the binding energy reaches the minimum at $R_{\epsilon=1} = 2.45$ Å in the gas phase and $R_{\epsilon=2} = 2.60$ Å in the low dielectric. The $R_{\epsilon=2}$ value is shifted to a larger Na−Cl distance relative to $R_{\epsilon=1}$, since the effect of the dielectric is to preferentially stabilize the oppositely charged ions as they separate.

The middle and bottom parts of Figure 2 show that the electrostatic low-dielectric "solvation" energy gradient has a minimum at $R = 3.60$ Å, whereas the corresponding solvation energy decreases monotonically. Both the solvation energy and gradient are continuous and differentiable (smooth). The gradient of the solvation energy, dE_{pol}/dE, shows some small but stable oscillatory features that represent the smooth appearance of surface elements as the ions separate. Note that the gradient oscillations are almost undetectable in the corresponding energy curve E_{pol}.

3.2. Geometry Optimizations of Biological Phosphorus Compounds. The purpose of this subsection is to test the implementation of the MNDO/d-smooth COSMO for stable geometry optimizations of anionic and neutral biological phosphorus compounds. To accomplish this, different components of the solvation energy are examined on the gas phase and solution potential energy surface and the expected trends are discussed.

The total solvation energy, ΔE_{sol}^{tot} is given as

$$\Delta E_{sol}^{tot} = (E_0[R] + \Delta E_{sol}[R]) - E_0[R_0] = E[R] - E_0[R_0] \quad (59)$$

where E_0 and E refer to the gas- and solution-phase total energies, respectively, at the electronically relaxed geometry indicated in brackets, $\Delta E_{sol} = E - E_0$ is the corresponding solvation energy contribution at a particular fixed geometry (hence, neglecting structural relation). Shown in brackets are the geometries optimized on the gas-phase PES (R_0) or solution-phase PES (R). For the systems studied in the present work, the computational cost of single-point energy and gradient calculation with MNDO/d-SCOSMO increased relative to the gas phase MNDO/d by a factor of 4−5. For anions, the self-consistent field procedure was considerably more stable with SCOSMO than in the gas phase. Geometry optimizations in the gas phase and in solution took similar number of steps.

Figure 3 illustrates the thermodynamic process of solvation which is decomposed into solvent induced electronic charge redistribution (vertical processes) and structural relaxation (horizontal processes)

$$\Delta E_{sol}^{tot} = \Delta E_{sol}[R_0] + \Delta E_{rel} = \Delta E_{str} + \Delta E_{sol}[R] \quad (60)$$

where the difference energy terms are defined as $\Delta E_{sol}[R_0] = E[R_0] - E_0[R_0]$, $\Delta E_{rel} = E[R] - E[R_0]$, $\Delta E_{str} = E_0[R] - E_0[R_0]$, and $\Delta E_{sol}[R] = E[R] - E_0[R]$. The quantities $\Delta E_{sol}(R_0)$ and $\Delta E_{sol}(R)$ are the solvation energies at the gas- and solution-phase geometries, respectively. The relaxation energy ΔE_{rel} represents the energy gain associated with the structural relaxation from R_0 to R in solution, and ΔE_{str} is the unfavorable strain energy introduced by perturbing the structure from R_0 to R in the gas phase.

The neutral and anionic trivalent (metaphosphate), tetravalent (phosphate) and pentavalent (phosphorane) phosphorus compounds are models for the reactants and products involved in the biological phosphoryl transfer and phosphate hydrolysis reactions.[54,55] The gas- and solution-phase energies as well as the RMS deviations between the gas- and solution-phase structures of the model compounds are summarized in Table 1. The total solvation energy, ΔE_{sol}^{tot}, for the molecules listed in Table 1 ranges from −9.9 to −25.4 kcal/mol for the neutral forms and from −75.0 to −81.3 kcal/mol for the monoanionic forms. Among both neutral and monoanionic species, the most favorable total solvation energy occurs for the smallest metaphosphate, and the least favorable solvation energy occurs for the largest acyclic phosphorane.

The stabilization due to the solvent-induced structural relaxation (ΔE_{rel}) generally increases with the increasing degrees of freedom and the total charge of the molecule. For the neutral phosphorus compounds, ΔE_{rel} ranges up to −1.4 kcal/mol and makes up typically less than 5% of the total solvation energy. For the corresponding anionic forms, ΔE_{rel} ranges up to −2.6 kcal/mol. Both acyclic and cyclic phosphorane anions are subject to appreciable energy stabilization (about 2.5 kcal/mol) due to the solvent-induced structural relaxation. Dimethyl phosphate has gauche−gauche (g−g) and gauche−trans (g−t) conformations that are close in energies. The combined MNDO/d-SCOSMO method yields lower solution-phase energy and a slightly smaller solvation energy for the g−g conformation, in agreement with previous studies using ab initio methods.[56,57] The quantities ΔE_{str} and ΔE_{rel} are similar in magnitude, but of opposite sign, which is plausible since they correspond to reverse structural relaxation processes. The range of ΔE_{rel} and ΔE_{str} values in Table 1 suggests that inclusion of solvent-induced structural relaxation may be an important factor, on the order of a few kcal/mol, for anionic biological phosphorus compounds.

Solvent-induced structural perturbations for the molecules listed in Table 1 can result in appreciable changes in torsion angles involving hydrogen atoms (data not shown). However, the heavy atom RMS deviations are relatively small, ranging from 0.02 to 0.1 Å. The largest structural relaxation occurs in acyclic phosphorane monoanion. In particular, solvation opens the axial O−P−O bond angle by about 5° and moves the hydrogen atom (from the O−H bond) by about 0.3 Å further away from the closest axial oxygen atom. These results emphasize the importance of solvent-induced structural relaxations for anionic phosphoranes that form transition states and intermediates in the biological transphosphorylation reactions.

Recently, there has been a comprehensive study[57] of the structure and stability of biological metaphosphate, phosphate, and phosphorane compounds in the gas phase and in solution based on density-functional calculations and continuum solvation

TABLE 1: Energies and Structural Deviations of Biological Phosphorus Compounds in the Gas Phase and in Solution[a]

molecule	ΔE_{sol}^{tot}	$\Delta E_{sol}(R_0)$	ΔE_{rel}	ΔE_{str}	$\Delta E_{sol}(R)$	RMSD
metaphosphates						
P(O)(O)(OH)	−25.4	−24.8	−1.4	0.6	−26.2	0.028
$P(O)(O)(O)^-$	−81.3	−81.3	−0.0	0.0	−81.3	0.018
acyclic phosphates						
$P(O)(OH)(OCH_3)(OCH_3)$	−16.9	−16.0	−0.9	1.1	−18.0	0.049
$P(O)(O)(OCH_3)(OCH_3)^-$ (g−g)	−79.3	−77.0	−2.3	2.8	−82.1	0.043
$P(O)(O)(OCH_3)(OCH_3)^-$ (g−t)	−79.8	−77.7	−2.1	2.3	−82.1	0.030
cyclic phosphates						
$P(O)(OH)(-OCH_2CH_2O-)$	−14.8	−14.2	−0.7	0.6	−15.5	0.049
$P(O)(O)(-OCH_2CH_2O-)^-$	−78.9	−77.2	−1.8	2.1	−81.0	0.019
acyclic phosphoranes						
$P(OH)(OH)(OCH_3)(OCH_3)(OCH_3)$	−9.9	−9.6	−0.3	0.3	−10.2	0.025
$P(O)(OH)(OCH_3)(OCH_3)(OCH_3)^-$	−75.0	−72.4	−2.6	2.4	−77.4	0.100
cyclic phosphoranes						
$P(OH)(OH)(-OCH_2CH_2O-)(OCH_3)$	−13.6	−13.1	−0.5	0.7	−14.3	0.032
$P(O)(OH)(-OCH_2CH_2O-)(OCH_3)^-$	−78.8	−76.3	−2.5	2.3	−81.1	0.033

[a] For definitions of different components of the total solvation energy, ΔE_{sol}^{tot}, see Figure 3 and eqs 59 and 60. All energetic quantities are in kcal/mol. Heavy-atom (non-hydrogen) root-mean-square deviations (RMSD) between structures optimized in the gas phase and in solution are in Å.

methods. The magnitudes of the solvation energies reported here are slightly larger than those given previously,[57] likely due to the use of Bondi radii[50] that were not optimized for use with the present smooth COSMO method. Tables S1 and S2 in the Supporting Information show the performance of the MNDO/d-SCOSMO method with Bondi radii and neglecting nonpolar contributions in calculating solvation energies of ions. Solvation energies with Bondi radii are on average slightly too negative for cations and especially for the anions. Scaling the Bondi radii by a factor of 1.1 improves the agreement with experimental solvation energies significantly. A full re-parametrization of solvation radii for use in the MNDO/d-SCOSMO method is forthcoming.

It is a worthwhile endeavor to further develop new semiempirical Hamiltonian models[58] in conjunction with implicit solvation methods in order to accurately predict solvation free energies and related properties of biological phosphorus compounds, in particular pK_a shifts of phosphoranes.[59,60] The method developed here represents an important step toward this goal. It remains to further calibrate and parametrize semiempirical quantum methods for obtaining accurate gas-phase proton affinities (an important step in the thermodynamic cycle to obtain pK_a values),[61] in concert with the improvement of the smooth COSMO solvation model (parametrization of the radii and nonelectrostatic terms) to arrive at a reliable tool for the prediction of pK_a shifts for biological phosphates and phosphoranes.

3.3. Potential Energy Surface for Phosphoryl Transfer of Methyl Phosphate. In this subsection, the smooth COSMO solvation method is applied to the dissociative mechanism of phosphoryl transfer in methyl phosphate and compared with results obtained from the conventional COSMO solvation method as implemented in MNDO97.[49] The focus here is on the numerical stability of the smooth COSMO method relative to the conventional COSMO method.

In this reaction, the phosphoryl group of methyl phosphate ($[CH_3O-PO_3]^{2-}$) dissociates as a metaphosphate anion (PO_3^-), leaving behind a methoxide anion (CH_3O^-). This reaction is the first step in the dissociative phosphoryl transfer pathway, a systematic theoretical study of which has been described in detail by others.[62] Since this step of the reaction can be described with a dianionic D_N type mechanism[63] where the departing groups are both monoanions, Coulomb repulsion preferentially stabilizes the dissociated species. Solvation effects to a large extent counterbalance the Coulomb effects by shielding the inter-ionic Coulomb repulsion and preferentially stabilizing the dianionic reactant and transition state complexes relative to the dissociated monoanionic products.

TABLE 2: Comparison of Relative Energy Values for Stationary Points Along the Dissociative Pathway for Phosphoryl Transfer of Methyl Phosphate in Solution as a Function of Discretization Level[a]

	reactant				transition state			
N_{se}	ΔE_{str}	$E_0(R)$	$\Delta E_{sol}(R)$	$E(R)$	ΔE_{str}	$E_0(R^{\ddagger})$	$\Delta E_{sol}(R^{\ddagger})$	$E(R^{\ddagger})$
26	4.46	65.42	−96.95	−31.53	2.88	73.49	−71.48	2.01
50	4.67	65.64	−96.94	−31.30	5.20	75.81	−74.32	1.49
110	4.62	65.58	−96.81	−31.23	8.15	78.76	−77.50	1.26
194	4.58	65.54	−96.71	−31.17	8.26	78.86	−77.73	1.14
302	4.63	65.59	−96.80	−31.22	7.95	78.56	−77.45	1.10
434	4.63	65.59	−96.81	−31.21	8.11	78.72	−77.63	1.09
590	4.63	65.59	−96.80	−31.21	8.13	78.74	−77.66	1.08
770	4.63	65.59	−96.81	−31.22	8.14	78.75	−77.68	1.07
974	4.63	65.59	−96.81	−31.22	8.16	78.76	−77.69	1.08
1202	4.64	65.60	−96.82	−31.22	8.18	78.79	−77.72	1.07

[a] Relative energy values (kcal/mol) with respect to infinitely separated species. Shown are the total energy in solution $E(R)$, and the solvation energy components (see Figure 3, and eqs 59 and 60).

3.3.1. Convergence of Stationary Points with Discretization Level. Table 2 shows the convergence of the unconstrained reactant and transition state energy values relative to the infinitely separated monoanionic species for stationary points optimized in solution as a function of the discretization level of cavity surface in the smooth COSMO method. Both the relative reactant and transition state energies converge from above. At very high discretization levels, the relative energy values agree to within 0.01 and 0.03 kcal/mol for the reactant and transition state, respectively. The geometries of the optimized stationary points converge much more rapidly with surface discretization (data not shown) such that the observed differences in relative energies arise mainly from the solvation energy terms. This is evident in Table 2 by inspection of the gas-phase relative energy values (E_0) evaluated at the solution-phase stationary points (either the reactant minimum R or the transition state $R^{\ddagger}$). The $E_0(R)$ and $E_0(R^{\ddagger})$ values range from 65.42 to 65.60 kcal/mol and from 73.49 to 78.79 kcal/mol, respectively. The greater range in the $E_0(R^{\ddagger})$ values (5.30 kcal/mol) relative to the range of $E_0(R)$ values (0.18 kcal/mol) reflects the relatively loose character of the transition state that is more

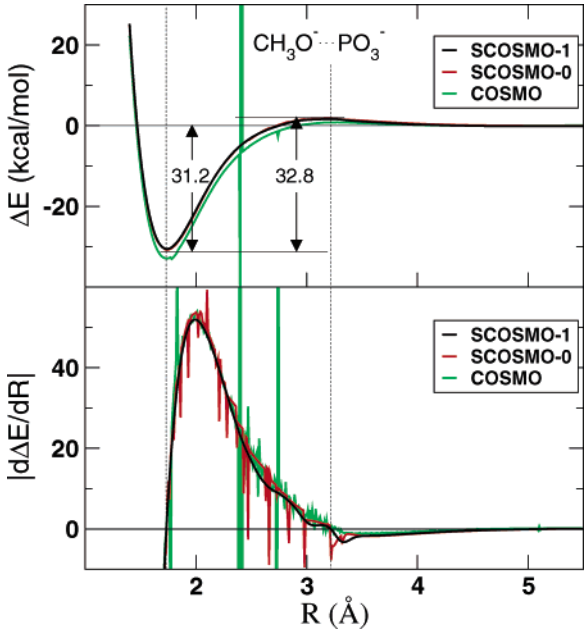

Figure 4. Reaction energy profile of the dissociative pathway for phosphoryl transfer of methyl phosphate in solution using the MNDO/d-smooth COSMO model (SCOSMO-1), the MNDO/d-smooth COSMO model without surface element switching (SCOSMO-0), and the conventional COSMO method as implemented in MNDO97 (COSMO). Shown are the binding energy values with respect to the infinitely separated products (ΔE, top) in kcal/mol and the associated gradient norm with respect to the reaction coordinate ($|d\Delta E/dR|$, bottom) in kcal/mol/Å. The reaction coordinate is defined as the $CH_3O^-\cdots PO_3^-$ distance (R).

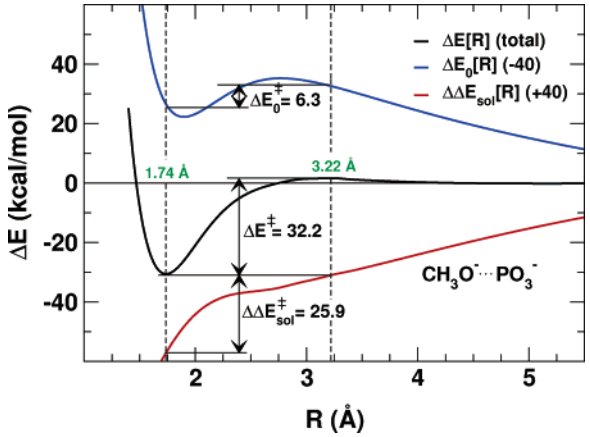

Figure 5. Reaction energy profile of the dissociative pathway for phosphoryl transfer of methyl phosphate in the gas phase and in solution using the MNDO/d-smooth COSMO model. Shown are the binding energy values (black curve), in addition to the shifted gas-phase (blue curve) and solvation energy components (red curve) along the reaction coordinate defined as the $CH_3O^-\cdots PO_3^-$ distance (R). In order for all the energy curves to appear on the same scale, the gas-phase energy (ΔE_0) and solvation energy ($\Delta\Delta E_{sol}$) components are shifted by -40 and $+40$ kcal/mol, respectively.

sensitive to changes in the solvation energy than the tighter reactant minimum.

The convergence behavior of the energies with respect to surface discretization is a feature that is well-behaved in the present smooth COSMO solvation method (typically much better than that of other boundary element solvation methods thus far tested). All boundary element solvation models, as well as the finite difference Poisson or Poisson−Boltzmann methods, have issues with regard to their convergence behavior that is not trivial to characterize. Ultimately, however, the accuracy of an applied solvation method depends, in part, on the parameters that are obtained by fitting to experimental or theoretical results at a particular discretization level.

3.3.2. Comparison of PES with Conventional and Smooth Solvation Models. Figure 4 compares the energy and gradient along the reaction coordinate $CH_3O^-\cdots PO_3^-$ for the phosphoryl transfer in methyl phosphate obtained from the smooth COSMO, the smooth COSMO without switching function (conventional COSMO with Gaussian surface elements), and the conventional COSMO methods as implemented in MNDO97.[49] The energy profile generated with the smooth COSMO method has stationary points at the reaction coordinate values of 1.74 Å (dianionic reactant) and 3.22 Å (transition state), respectively, that are coincident with the points where the corresponding gradients have zero values. The conventional COSMO method (green curve, designated "COSMO" in Figure 4) results in singularities in both binding energy and gradient curves. This is caused by the numerical instability in the matrix elements due to the point-charge approximation[14] and more importantly, the fluctuation in the dimensionality of the interaction matrixes due to the appearance or disappearance of surface elements.[15] These factors affect both the solvent response as well as the solvent-induced electronic polarization energy. Inclusion of Gaussian functions provides numerical stability for the matrix elements. As a result, the energy curve appears to be continuous although with little nonsmoothness; however, singularities persist in the gradients (red curve, designated "SCOSMO-0" in Figure 4). Finally, the additional inclusion of a smooth switching function circumvents the problem with regard to the appearance or disappearance of surface elements. Both energy and gradient values are smooth with respect to geometrical changes (black curve, designated "SCOSMO-1" in Figure 4). It is worthwhile to note that, out of the 411 geometry optimizations making up the energy curves, all 411 of the conventional COSMO calculations and 380 of SCOSMO-0 calculations failed to reach the convergence criteria of 0.01 kcal/mol/Å on the gradient norm. However, all calculations performed with the SCOSMO-1 method successfully met the convergence criteria.

Comparison between energy and gradient curves in Figure 4 reveals that "small" nonsmoothness in the energy profile (such as in the SCOSMO-0 curve) can develop into singularity problems in the gradients leading to numerical instability and failure in the optimization procedures. The problem is typically exacerbated with increasing degrees of freedom,[33] especially in large-scale geometry optimizations and transition state searches that utilize linear-scaling techniques.[26] Most recently, alternative strategies have been explored to address the discontinuity problem in boundary element solvation methods.[33,64]

3.3.3. Effect of Solvation on the PES. Figure 5 compares the energy profile of the phosphoryl transfer during the methyl phosphate reaction in solution and in the gas phase. The dianionic reactant complex is a stable energy minimum in both cases. In the gas phase, the reactant complex is less stable than the dissociated monoanionic metaphosphate and methoxide, whereas in solution, it is more stable by about 30 kcal/mol due to more favorable solvation compared with the dissociated product. The same effect causes a shift of the reactant minimum from 1.90 Å in the gas phase to 1.74 Å in solution. At the transition state ($R = 3.22$ Å), the solution-phase activation energy barrier relative to the dianionic reactant complex is 32.2 kcal/mol, in reasonable quantitative agreement with the experimental value of 37 ± 3 kcal/mol.[65] This energy barrier can be decomposed into the gas-phase energy ($\Delta E_0^{\ddagger}$) of 6.3 kcal/mol, and the solvation contribution ($\Delta\Delta E_{sol}^{\ddagger}$) of 25.9 kcal/mol. The

transition state is significantly destabilized by solvation relative to the reactant and product states due to the larger effective radius of the associated dianionic complex, causing a shift from 2.77 Å in the gas phase to 3.22 Å in solution. In accord with the Hammond postulate,[66] the preferential reactant stabilization by solvent increases the forward activation barrier and causes the transition state to be shifted toward the products. The result is a very late transition state, consistent with a dissociative mechanism, with a higher forward activation barrier than would occur in a low dielectric environment.

These results demonstrate the stability of the MNDO/d-SCOSMO method and illustrate the general effect of solvation on phosphoryl transfer reactions that involve association/dissociation of like-charged ionic species. The activation barrier calculated by the MNDO/d-SCOSMO method is in reasonable agreement with the experimental value. It is encouraging that realistic results are obtained with the use of standard solvation parameters and a robust d-orbital quantum model, neither of which has been specifically parametrized to obtain accurate energies for biochemical reactions. A more comprehensive study of this reaction would involve inclusion of explicit water molecules[62] and QM/MM simulation with specific reaction parameters for phosphoryl transfer reactions. Previous tests of the MNDO/d method for dianionic reaction mechanisms for phosphate diesters, however, agree reasonably well with density-functional results,[58] as do application results from hybrid QM/MM simulations.[67,68] The concurrent development of semiempirical quantum methods and improved parameters for implicit and QM/MM solvation methods is an area of current effort that can greatly benefit from the results of the present work.

4. Conclusion

The present paper describes the implementation of analytic gradients for the smooth COSMO method and integration within the d-orbital semiempirical framework.[49] A correction for multiple-atom switching is also introduced into the original smooth COSMO formulation.

The combined MNDO/d-SCOSMO method has been tested on the potential energy curve for separating Na^+ and Cl^- ions to demonstrate the smoothness in the solvation energy and gradient along the internuclear axis, and coincidence of the energy minimum with the zero-gradient norm. Tests of the method on the energies and structures for a series of neutral and charged metaphosphate, phosphate, and phosphorane compounds were performed. Solvent-induced structural relaxation was typically less than 1.4 kcal/mol for neutral compounds and 2.6 kcal/mol for the anionic compounds. The overall effects on the structure were moderate.

Application to the energy and gradient profile for the dissociative phosphoryl transfer of methyl phosphate reaction illustrates the stability of the method, and the failure of other nonsmooth solvation models to produce smooth profiles. The convergence of the smooth COSMO method with respect to discretization level is shown to be well behaved both for energy minima and transition states. Tests on the dianionic dissociation of methyl phosphate provide insight into the role of solvation in reactions that involve the dissociation (or association) of monoanions to form a dianionic transition state complex. The magnitude of the solvation energy is observed to decrease monotonically as a function of the dissociative reaction coordinate. The overall effect of solvation is to preferentially stabilize the associated dianionic complexes relative to the dissociated reactants and increase the activation barrier relative to that of the gas-phase reaction in accord with the Hammond postulate.

The current work demonstrates the numerical stability of the smooth COSMO method in performing solution-phase quantum mechanical geometry optimizations and transition state searches for biologically important molecules, for which the previous COSMO implementation fails. The results of the present work allow application of the MNDO/d-SCOSMO method to model chemical reactions in solution.[51] One of the major advantages of the MNDO/d-SCOSMO method is the significantly lower computational cost relative to QM/MM simulations with explicit solvent. Although further effort including radii optimization is warranted, the presented results are encouraging and may assist in the design of new implicit solvent models that provide increased accuracy and transferability for biological reactions.

Acknowledgment. D.Y. is grateful for financial support provided by the National Institutes of Health (Grant GM62248) and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement Number DAAD19-01-2-0014. The content does not necessarily reflect the position or the policy of the government and no official endorsement should be inferred. Computational resources were provided by the Minnesota Supercomputing Institute.

Supporting Information Available: Solvation energies for the MNDO/d smooth cosmo method using Bondi radii are presented for a set of cations and anions, and compared to experiment. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(2) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(3) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1600.
(4) Makarov, V.; Pettitt, B. M.; Feig, M. *Acc. Chem. Res.* **2002**, *35*, 376−384.
(5) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027−2094.
(6) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161−2200.
(7) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(8) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(9) Im, W.; Beglov, D.; Roux, B. *Comput. Phys. Commun.* **1998**, *111*, 59−75.
(10) Gregory D. Hawkins, C. C.; Truhlar, D. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(11) Im, W.; Lee, M. S.; Brooks, C. L., III. *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(12) Miertuš, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117−129.
(13) Rashin, A. A. *J. Phys. Chem.* **1990**, *94*, 1725−1733.
(14) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.
(15) York, D. M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 11060−11079.
(16) Yang, W.; Pérez-Jordá, J. M. Linear scaling methods for electronic structure calculation. In *Encyclopedia of Computational Chemistry*; von Schleyer, P. R., Ed.; John Wiley and Sons: New York, 1998.
(17) Goedecker, S. *Rev. Mod. Phys.* **1999**, *71*, 1085−1123.
(18) Goedecker, S.; Scuseria, G. E. *IEEE Comput. Sci. Eng.* **2003**, *5*, 14−21.
(19) York, D. M.; Lee, T.-S.; Yang, W. *Chem. Phys. Lett.* **1996**, *263*, 297−306.
(20) Pomelli, C. S.; Tomasi, J. *J. Mol. Struct. (THEOCHEM)* **2001**, *537*, 97−105.
(21) Rega, N.; Cossi, M.; Barone, V. *Chem. Phys. Lett.* **1998**, *293*, 221−229.
(22) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.
(23) Khandogin, J.; York, D. M. *J. Phys. Chem. B* **2002**, *106*, 7693−7703.
(24) Khandogin, J.; Musier-Forsyth, K.; York, D. M. *J. Mol. Biol.* **2003**, *330*, 993−1004.
(25) Khandogin, J.; York, D. M. *Proteins* **2004**, *56*, 724−737.

(26) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177−2186.
(27) Cortis, C.; Friesner, R. A. *J. Comput. Chem.* **1997**, *18*, 1570−1590.
(28) Cortis, C.; Friesner, R. *J. Comput. Chem.* **1997**, *18*, 1591−1608.
(29) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.
(30) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224−2229.
(31) Cossi, M.; Mennucci, B.; Cammi, R. *J. Comput. Chem.* **1996**, *17*, 57−73.
(32) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2002**, *117*, 43−54.
(33) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449−1462.
(34) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.
(35) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.
(36) Truong, T. N.; Nguyen, U. N.; Stefanovich, E. V. *Int. J. Quantum Chem.* **1996**, *30*, 403−410.
(37) Stefanovich, E. V.; Truong, T. N. *J. Phys. Chem. B* **1998**, *102*, 3018−3022.
(38) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(39) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340−366.
(40) Chipman, D. M. *J. Chem. Phys.* **1997**, *106*, 10194−10206.
(41) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622−6633.
(42) Zhan, C.; Bentley, J.; Chipman, D. M. *J. Chem. Phys.* **1998**, *108*, 177−192.
(43) Stroud, A. H. *Approximate Calculation of Multiple Integrals*; Prentice Hall: Englewood Cliffs, NJ, 1971.
(44) Lebedev, V. I. *Sibirsk. Mater. Ž* **1977**, *18*, 132−142.
(45) Delley, B. *J. Comput. Chem.* **1996**, *17*, 1152−1155.
(46) Floris, F. M.; Tomasi, J.; Pascual-Ahuir, J. L. *J. Comput. Chem.* **1991**, *12*, 784−791.
(47) Floris, F. M.; Selmi, M.; Tani, A.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 6353−6365.
(48) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, *46*, 89−104.
(49) Thiel, W. *MNDO97*, version 5.0; University of Zurich: Zurich, Switzerland, 1998.
(50) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441−451.
(51) Gregersen, B. A.; Khandogin, J.; Thiel, W.; York, D. M. *J. Phys. Chem. B*, in press.
(52) Bontha, J. R.; Pintauro, P. N. *J. Phys. Chem.* **1992**, *96*, 7778−7782.
(53) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(54) Zhou, D.-M.; Taira, K. *Chem. Rev.* **1998**, *98*, 991−1026.
(55) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 432−450.
(56) Florián, J.; Štrajbl, M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *120*, 7959−7966.
(57) Range, K.; McGrath, M. J.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 1654−1665.
(58) Lopez, X.; York, D. M. *Theor. Chem. Acc.* **2003**, *109*, 149−159.
(59) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010−5018.
(60) Davies, J.; Doltsinis, N.; Kirby, A.; Roussev, C.; Sprik, M. *J. Am. Chem. Soc.* **2002**, *124*, 6594−6599.
(61) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421−6427.
(62) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(63) Guthrie, R. D.; Jencks, W. P. *Acc. Chem. Res.* **1989**, *22*, 343−349.
(64) Pomelli, C. S. *J. Comput. Chem.* **2004**, *25*, 1532−1541.
(65) Guthrie, J. P. *J. Am. Chem. Soc.* **1977**, *99*, 3991−4001.
(66) Hammond, G. S. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(67) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.
(68) Gregersen, B. A.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.